

A Study on a Hybrid Software Quality Management Framework Based on Collaboration between Artificial Intelligence and Human Experts

Eun-Sook Cho*

*Professor, Dept. of Software Engineering, Seoil University, Seoul, Korea

[Abstract]

The adoption of artificial intelligence (AI) has improved automation and efficiency in software quality management; however, complex domain interpretation and ethical judgment still require human expertise. Existing approaches tend to focus on either AI-driven automation or human-centered processes, with limited integration of both. This paper proposes a hybrid software quality management framework that combines AI automation with human expert judgment to enhance reliability and consistency. The framework is based on four principles—role optimization, interactive feedback, multi-layered quality evaluation, and continuous improvement—and is aligned with ISO/IEC 25010 quality characteristics. Experimental evaluations conducted in smart manufacturing, educational AI (EduTech), and cloud service quality management demonstrate that the proposed hybrid approach effectively improves the speed and reliability of quality management. These results indicate that the framework can serve as a practical model for AI-human collaborative quality management in industrial applications.

▶ **Key words:** Hybrid Software Quality Management, AI-Human Collaboration, AI Automation, Quality Reliability and Consistency ISO/IEC 25010, Industrial Applications

[요 약]

인공지능(AI)의 도입은 소프트웨어 품질관리의 자동화와 효율성을 향상시켰으나, 복잡한 도메인 해석과 윤리적 판단에는 여전히 인간 전문가의 개입이 요구된다. 기존 연구들은 AI 중심 자동화 또는 인간 중심 품질관리 방식에 치우쳐, 두 접근을 통합한 연구는 제한적이었다. 이에 본 논문에서는 AI 자동화와 인간 전문가의 판단을 결합한 하이브리드 소프트웨어 품질관리 프레임워크를 제안하여 품질관리의 신뢰성과 일관성을 향상시키고자 하였다. 제안된 프레임워크는 역할 최적화, 상호 피드백, 다층 품질평가, 지속적 개선의 네 가지 원칙을 기반으로 하며 ISO/IEC 25010의 품질 특성과 연계된다. 스마트 제조, 에듀테크, 클라우드 서비스 도메인에서 수행한 실험 결과, 제안된 하이브리드 접근은 품질관리의 처리 속도와 신뢰성을 효과적으로 개선함을 확인하였다. 본 연구는 AI-인간 협업 기반 품질관리의 실무적 적용 가능성을 제시한다.

▶ **주제어:** 하이브리드 소프트웨어 품질관리, 인공지능-인간 협업, AI 자동화, 품질 신뢰성 및 일관성, ISO/IEC 25010, 산업 응용

• First Author: Eun-Sook Cho, Corresponding Author: Eun-Sook Cho
*Eun-Sook Cho (escho@seoil.ac.kr), Dept. of Software Engineering, Seoil University
• Received: 2025. 11. 10, Revised: 2025. 12. 19, Accepted: 2025. 12. 19.

I. Introduction

소프트웨어 품질관리는 제품의 신뢰성, 안정성, 성능, 보안성을 확보하기 위한 핵심 활동으로, 대규모 시스템 개발이 보편화된 오늘날에는 품질관리의 자동화·지속화가 중요한 과제가 되었다[1-3].

초기 품질관리 연구는 주로 인간 전문가의 경험적 관리에 의존했으나, 데이터 규모의 폭증과 복잡한 요구사항의 증가로 인해 한계에 봉착하였다. 이에 따라 AI 기반 자동화 기술이 품질관리의 효율성을 높이는 대안으로 등장하였다. 특히 2023~2025년 사이의 연구[4-8]는 테스트 자동화, 결함 탐지, 코드 품질분석 등에서 AI의 역할을 실증적으로 확장하였다. 그러나 AI 단독 접근은 편향(Bias), 책임 소재, 윤리적 판단 부재 등 고차원적 문제를 완전히 해결하지 못하였다. 이 한계는 결국 인간 전문가의 판단을 결합한 하이브리드 품질관리(Hybrid Quality Management) 연구로 이어졌다[9-12]. 즉, 최근의 AI 연구들이 제시한 자동화 성과 위에, 2019년부터 2024년까지의 협업 연구들이 결합되며 새로운 방향성이 형성된 것이다.

본 연구는 이러한 학문적 흐름을 통합하여, AI의 속도·일관성과 인간의 창의·윤리적 판단을 결합한 협업 기반의 하이브리드 품질관리 프레임워크를 제안한다.

이 프레임워크는 반복적·표준화된 작업은 AI가 자동화하고, 맥락적·윤리적 판단이 필요한 단계에서는 인간 전문가가 개입하여 상호 피드백을 주고받는 구조로 설계된다.

이를 통해 기존 연구가 분리적으로 접근했던 효율성과 신뢰성(Human Oversight)을 동시에 달성하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 협업 기반 품질관리와 관련된 선행연구를 분석하고, 3장에서는 제안 프레임워크의 설계 원칙 및 구조를 제시한다. 4장에서는 세 산업 도메인(스마트 제조, EduTech, 클라우드 서비스)을 대상으로 실험을 수행하여 제안 모델의 실효성을 검증하고, 5장에서 연구의 결론 및 향후 연구 과제를 제시한다.

II. Preliminaries

1. Related works

본 장에서는 인간-AI 협업(Human-AI Collaboration, 이하 HAC)에 기반한 품질관리 연구를 검토하고, AI 단독 접근, 인간 중심 접근, 하이브리드 접근의 차이점과 보완점을 제시한다.

1.1 Human-centered Quality Evaluation

인간 중심 품질관리 연구[4-6,10]는 전문가의 경험과 도메인 지식에 근거한 평가 체계를 강조하였다.

이 접근은 복잡한 시스템의 맥락적 이해에는 강점을 가지지만, 데이터 처리 효율과 일관성 면에서 한계가 있다.

따라서 품질관리에서 인간과 AI의 역할 분담 및 통합의 필요성이 대두되었다.

1.2 AI-driven Quality Automation

AI 기반 품질관리 연구[6-9]는 주로 테스트 자동화, 결함 탐지, 코드 생성 등 반복 작업의 효율화에 초점을 두었다. 예를 들어 Wu et al.(2025)[6]은 생성형 AI와 인간의 협업을 통해 업무 효율성과 산출물 품질이 향상됨을 보였으나, 윤리적 판단·설명가능성 영역에서는 여전히 인간 개입이 필요함을 지적하였다. 이러한 연구들은 품질관리 프로세스에서 자동화의 성취는 입증했지만, 책임과 해석의 영역을 충분히 고려하지 못했다.

1.3 Hybrid and Collaborative Approaches

최근 연구들은 AI 자동화의 강점과 인간 판단의 책임성을 결합한 협업형(Hybrid) 품질관리 프레임워크를 제시하기 시작했다[8-9].

Fragiadakis et al.(2024)[8]은 HAC 시스템을 평가하는 방법론을 제안하며, 협업 품질의 핵심요소를 다음 네 가지로 정의했다: ① 상호 신뢰(Trust), ② 역할 명확성(Role Clarity), ③ 정보 공유(Information Transparency), ④ 공동 의사결정(Joint Decision Making).

Table 1. Comparison of Quality Management Approaches

Type	Key Characteristics	Advantages	Limitations
AI-driven	Automation-oriented	High speed Consistency	Lacks ethical consideration; potential bias
Human-centered	Experience and judgment-based	Contextual understanding Flexibility	Limited efficiency and scalability
Hybrid (Collaborative)	Interactive feedback between AI and human experts	Combines accuracy with accountability	Absence of unified standard framework

이 네 요소는 본 연구의 프레임워크 설계에도 직접 반영된다. 요약하자면, [Table 1]에 제시된 것처럼 기존 연구들의 한계는 다음 세 가지로 정리된다. 첫째, 협업의 정의 및 평가 기준 불명확하다는 것이다. 둘째, 기술적 품질(성능·보안)과 인적 품질(윤리·신뢰) 간의 단절이 존재한다는 것이다. 셋째, 정량적·정성적 평가의 통합이 부족하다는 것이다.

이상의 선행 연구들은 모두 인간-AI 협업이 소프트웨어 품질 향상에 기여할 수 있음을 시사한다. 그러나 협업 품질의 평가 기준 표준화, 기술적 품질과 사용자 경험의 통합, 장기적 효과 및 윤리적 문제 등은 여전히 미해결 과제로 남아 있다. 따라서 본 연구는 이러한 한계를 극복하고, AI 자동화와 인간 전문가의 창의적·윤리적 판단을 결합한 하이브리드 소프트웨어 품질관리 프레임워크를 제안한다는 점에서 차별성을 가진다.

III. Designing Hybrid Software Quality Management's Framework

본 연구에서 제안하는 하이브리드 소프트웨어 품질관리 프레임워크의 개념적 구조는 [Fig. 1]과 같다. 이 구조는 AI의 자동화 능력과 인간 전문가의 창의적·윤리적 판단을 상호 보완적으로 결합하여 품질관리의 효율성과 신뢰성을 동시에 확보하는 것을 목표로 한다.

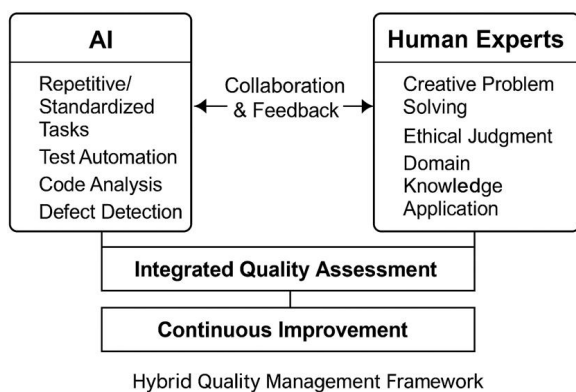


Fig. 1. Structure of the Hybrid Software Quality Management Framework

1. Principles for Framework Design

하이브리드 품질관리 프레임워크는 다음의 기본 원칙에 따라 설계된다.

- 역할 최적화: AI는 대량의 데이터 처리, 테스트 자동화, 코드 분석 등 반복적이고 규칙 기반의 품질관리

작업을 담당한다. 인간 전문가는 복잡한 문제 해결, 도메인 특화 판단, 윤리적·사회적 책임 등 고차원적 의사결정에 집중한다.

- 상호 피드백: AI가 산출한 결과물은 인간 전문가의 검토와 피드백을 거쳐 지속적으로 개선된다. 이 과정에서 AI는 인간의 피드백을 학습하여 품질관리의 정확성과 신뢰성을 높인다.
- 다층적 품질 평가: 기능적 품질(정확성, 효율성), 비기능적 품질(보안, 확장성, 사용자 경험), 윤리적 품질(공정성, 투명성, 책임성) 등 다양한 품질 요소를 통합적으로 평가한다.
- 지속적 개선: 품질관리 과정에서 발생하는 데이터와 경험을 체계적으로 축적·분석하여, 프레임워크 자체의 성능과 적응성을 지속적으로 향상시킨다.

이러한 4가지 설계 원칙은 ISO/IEC 25010 품질 모델의 핵심 속성(신뢰성, 성능 효율성, 보안성, 유용성)에 기반하여 도출되었다[13,14]. [Table 2]는 프레임워크 설계에 적용된 4개의 설계 원칙과 ISO/IEC 25010 품질모델의 품질 속성과의 연계 관계를 보여준다.

Table 2. Relationship between Design Principles and Quality Attributes

Design Principle	Corresponding Quality Attribute
Role Optimization	Performance Efficiency
Interactive Feedback	Reliability
Multi-layered Evaluation	Security and Reliability
Continuous Improvement	Maintainability and Usability

2. Phased Structure of the Framework

앞서 제시한 설계 기본 원칙을 바탕으로, 실제 소프트웨어 개발 현장에 적용 가능한 하이브리드 품질관리 프레임워크의 구체적 구조와 운영 프로세스를 제안한다.

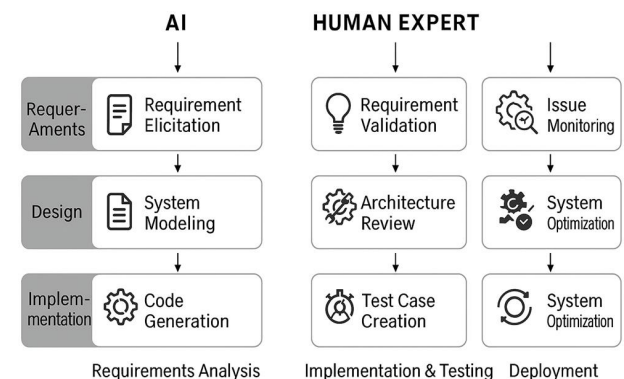


Fig. 2. Role Distribution of AI and Human Experts

제안된 프레임워크는 소프트웨어 개발 생명주기(SDLC)의 각 단계에서 AI와 인간 전문가의 역할을 명확히 구분·결합하도록 설계되었다. [Fig. 2]는 소프트웨어 개발 생명주기(SDLC)의 각 단계에서 AI와 인간 전문가가 담당하는 역할을 나타낸다. 요구분석 단계에서는 AI가 요구사항 문서화를 지원하고, 인간 전문가는 이를 검증한다. 설계, 구현·테스트, 배포 단계에서도 AI는 자동화된 작업을 수행하고, 인간 전문가는 검토와 최적화를 통해 품질을 확보한다. 이 구조를 통해 데이터 기반 객관성과 전문가의 경험적 통찰이 유기적으로 결합된다.

3. Interactive Feedback and Collaborative Process

AI와 인간 전문가 간의 협업이 효과적으로 작동하기 위해서는 단순한 역할 분담을 넘어, 상호 피드백과 공동 의사결정이 가능한 체계적 프로세스가 필요하다. 본 연구에서 제안하는 하이브리드 품질관리 프레임워크는 이를 위해 순환적 상호작용 구조를 중심으로 설계되었다.

이러한 순환적 피드백 구조와 공동 의사결정 과정을 도식화하면 [Fig. 3]과 같다.

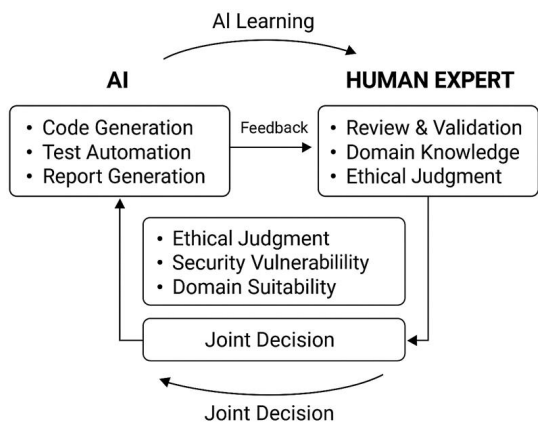


Fig. 3. Cyclical Feedback and Collaboration Process

AI가 산출한 결과물은 인간 전문가의 검토를 거쳐 보완되며, 이 피드백은 다시 AI 학습에 반영된다. 동시에 주요 품질 이슈는 AI의 데이터 분석과 인간 전문가의 맥락적 판단이 결합된 공동 의사결정을 통해 해결된다.

우선, AI는 반복적이고 표준화된 작업(예: 코드 생성, 테스트 자동화, 품질 리포트 작성)을 신속하게 수행한다. 그러나 AI가 산출한 결과물은 그대로 사용되지 않고, 반드시 인간 전문가의 심층적 검토와 보완 과정을 거친다. 이 과정에서 인간은 도메인 지식과 맥락적 판단, 윤리적 기준을 적용하여 결과물의 신뢰성과 적합성을 확보한다.

또한, 인간 전문가의 피드백은 일회성으로 소멸되지 않고 AI 학습 데이터로 반영된다. 이를 통해 AI는 반복 작업을 수행할수록 점차 정확성과 적응성을 높게 되며, 장기적으로 품질관리의 효율성과 신뢰성을 동시에 강화한다. 이러한 순환적 피드백 구조는 인간-기계 협업의 투명성과 신뢰성을 보장하는 핵심적 메커니즘이 된다.

나아가, 주요 품질 이슈(예를 들어 윤리적 판단, 보안 취약성, 도메인 적합성 등)는 단일 주체에 의해 결정되지 않고, AI의 데이터 분석·패턴 인식 역량과 인간 전문가의 책임 있는 판단과 맥락 해석을 결합한 공동 의사결정을 통해 해결된다. 이로써 복잡한 품질관리 문제에 대해 다각적 접근과 균형 잡힌 결론 도출이 가능하다.

결국, 상호 피드백 및 협업 프로세스는 본 프레임워크가 단순한 AI 보조 도구를 넘어, AI와 인간이 상호 학습하고 보완하는 동적 협업 체계임을 보여주며, 품질관리의 신뢰성과 지속 가능성을 뒷받침하는 핵심 기반이 된다.

4. Integrated Quality Assessment Framework

소프트웨어 품질관리는 단순히 기능적 오류를 검출하는 수준을 넘어, 시스템의 신뢰성과 사용자 경험, 나아가 윤리적 적합성까지 종합적으로 고려해야 한다. 제안한 프레임워크는 정량적 지표와 정성적 지표를 통합한 다차원적 평가 구조를 통해 품질관리의 효율성과 신뢰성을 동시에 고려할 수 있도록 설계되었다. [Fig. 4]는 정량적 평가와 정성적 평가가 통합 대시보드를 통해 연결되는 통합된 품질 평가 프레임워크 구조를 보여준다.

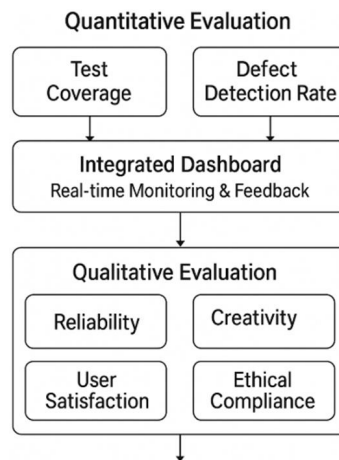


Fig. 4. Integrated Quality Assessment Framework

[Fig. 4]에서 제시된 통합 대시보드는 테스트 커버리지와 결합 탐지율 등 정량적 성능 지표와, 신뢰성·창의성·사용자 만족도·윤리 준수와 같은 정성적 품질 요소를 실시간

으로 연계하여 시각화한다. 이를 통해 AI가 제공하는 속도와 일관성 중심의 자동화 품질 정보와 인간 전문가가 제공하는 윤리적·맥락적 판단 기반의 신뢰성 정보가 상호 보완되며, 결과적으로 품질의 기술적 효율성과 인간적 가치가 조화를 이루는 균형적 품질관리 구조를 형성한다.

5. Continuous Improvement and Learning Processes

하이브리드 품질관리 프레임워크가 장기적으로 효과성을 유지하기 위해서는 일회성 관리가 아니라 지속적 개선 메커니즘이 필수적이다. 본 연구에서 제안하는 프레임워크는 품질관리 과정에서 발생하는 다양한 데이터와 피드백을 체계적으로 축적하고 이를 학습과 개선에 반영함으로써, 점진적으로 고도화되는 순환 구조를 지향한다.

[Fig. 5]는 제안된 프레임워크가 AI 자동화, 인간 검토, 피드백 수집, AI 학습 및 업데이트, 프로세스 최적화의 순환을 통해 지속적으로 개선되는 과정을 보여준다. 이러한 자기 진화적 구조(Self-evolving Framework)는 장기적 신뢰성과 적응성을 확보하는 핵심 기반이 된다.

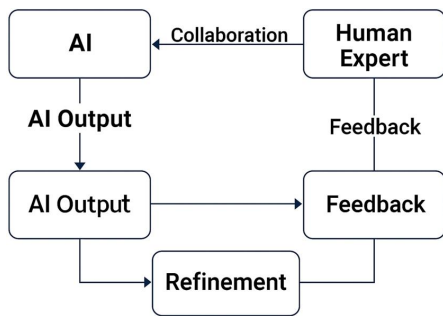


Fig. 5. Continuous Improvement Cycle

우선, 각 단계에서 수행된 품질관리 결과와 인간 전문가의 피드백은 데이터베이스에 기록된다. 이렇게 축적된 데이터는 단순 기록에 그치지 않고, AI 품질관리 도구의 재학습과 업데이트에 활용되어 자동화 성능을 지속적으로 향상시킨다. 이는 AI가 시간이 지남에 따라 더 높은 정확성과 적응성을 갖추도록 하는 핵심 기반이 된다.

또한, 현장에서 수집된 정성적·정량적 피드백은 프레임워크 자체의 운영 프로세스 개선에 직접 반영된다. 역할 분담의 조정, 평가 지표의 보완, 협업 프로세스의 정교화가 반복적으로 이루어짐으로써, 프레임워크는 고정된 체계가 아니라 변화하는 환경과 요구사항에 능동적으로 대응할 수 있는 적응적 품질관리 체계로 발전한다.

이러한 지속적 개선 과정은 단순히 기술적 성능 향상을 넘어, 인간과 AI 간의 협업 신뢰성을 강화하고 품질관리

전반의 투명성과 책임성을 제고하는 역할을 한다.

IV. Experiment & Evaluation

본 장에서는 제안된 하이브리드 소프트웨어 품질관리 프레임워크의 타당성과 실효성을 검증하기 위한 실험 설계 및 평가 과정을 설명하고, 정량적·정성적 분석을 통해 그 효과를 검증한다.

1. Experiment

1.1 Experiment Design

제안된 하이브리드 소프트웨어 품질관리 프레임워크의 유효성을 검증하기 위해 세 가지 산업 도메인—스마트 제조(Smart Manufacturing), 에듀테크(EduTech), 클라우드 서비스(Cloud Service Quality Management)—를 선정하였다. 이들 도메인은 모두 AI 자동화와 인간 전문가의 검증이 병행되는 품질관리 구조를 갖추고 있어, 본 연구의 프레임워크 적용 효과를 평가하기에 적합하다.

본 연구의 실험은 동일한 환경에서 수행되어 결과의 일관성과 재현성을 확보하였다. 모든 실험은 Intel Core i7 기반의 CPU(3.6GHz), 64GB RAM, NVIDIA RTX 3080 GPU(10GB VRAM)를 탑재한 워크스테이션 환경에서 수행되었으며, 운영체제는 Ubuntu 22.04 LTS를 사용하였다.

AI 분석 모델은 Python 3.10 환경에서 구현되었으며, 주요 라이브러리로는 Scikit-learn, TensorFlow, PyTorch, Pandas를 활용하였다. 데이터 전처리, 학습, 평가 과정은 동일한 파이프라인으로 구성하여 도메인 간 비교가 가능하도록 통제하였다. 품질관리 자동화 단계에서는 결함 탐지 및 분류를 위해 Random Forest, Gradient Boosting, 그리고 LSTM 기반 시계열 모델을 적용하였으며, 인간 전문가의 검증 단계에서는 AI가 생성한 분석 결과에 대해 오류 여부 및 맥락적 타당성을 평가하는 피드백 절차를 포함하였다. 모든 실험은 각 도메인별로 10회 반복 수행되었으며, 결과값은 평균값을 기준으로 분석하였다.

데이터는 공개된 산업용 오픈 데이터셋과 시뮬레이션 환경에서 수집되었다. 각 데이터셋은 평균 2만 건 이상의 관측치를 포함하였으며, 프레임워크 적용 전후를 비교하기 위해 반복 실험을 수행하였다. 통계적 유의성 검증은 paired t-test를 이용하였으며, 유의수준 $p < 0.05$ 에서 성능 차이를 평가하였다. 물론 인간 전문가의 판단과 피드백은 품질 향상에 기여하지만, 전문가 간 판단 편차와 주관성 문제가 발생할 수 있다. 본 연구에서는 이를 완화하기

위해 인간 전문가의 개입을 AI 분석 결과에 대한 검증 및 보완 역할로 한정하고, AI가 제공하는 신뢰도 지표와 표준화된 기준을 기반으로 의사결정이 이루어지도록 설계하였다. 또한 전문가 피드백은 반복 실험 과정에서 AI 모델에 재반영되어 판단 기준의 일관성을 점진적으로 확보하도록 구성하였다. 무엇보다 AI 학습 과정에서 발생할 수 있는 데이터 편향과 윤리적 오류를 완화하기 위해, 본 연구에서는 학습 데이터 분포 점검과 균형 조정을 수행하고, AI 분석 결과에 대해 인간 전문가가 최종 검증하는 구조를 적용하였다. 또한 품질관리 결과에 대한 최종 책임은 인간 전문가 및 조직에 귀속되도록 역할과 책임 범위를 명확히 정의하였다.

1.2 Industrial AI in Smart Manufacturing

스마트 제조 분야에서는 IoT 센서 데이터를 기반으로 공정 품질을 실시간 감시한다. 본 프레임워크는 전자 부품 제조사의 AI 품질관리 시스템에 적용되었으며, AI는 센서 데이터의 이상치 탐지와 결함 예측을 수행하고, 인간 전문가의 현장 경험을 바탕으로 결과를 검증하였다.

AI는 약 5만 건 이상의 생산 데이터를 분석하여 결함 탐지율과 불량 가능성을 예측하였으며, 인간 검토자는 예측 결과를 검증하고 오판된 사례를 피드백으로 제공하였다. 이 과정에서 결함 탐지 정확도는 89%에서 94%로 향상, 오류율은 14% 감소하였다. AI의 자동화 효율성과 인간의 경험적 판단이 조화를 이루어 공정 품질의 안정성과 예측 신뢰도가 개선되었다.

1.3 EduTech Platform Quality Management

에듀테크 도메인에서는 학습자 행동 로그(출석률, 과제 제출, 학습 성취도 등)를 기반으로 학습자의 성과를 예측하고 피드백을 제공하는 AI 시스템을 구축하였다. AI는 학습 패턴 데이터를 분석해 성취도를 예측하고, 인간 교사는 결과를 검토하여 학습자 맞춤형 피드백을 제공하였다.

프레임워크 적용 결과, 예측 정확도는 82%에서 90%로 향상, FPR(False Positive Rate)은 22% 감소하였다. AI의 정량적 분석과 인간 교사의 질적 판단이 결합되어 학습자의 참여도와 성취 품질이 모두 개선되었다.

1.4 Cloud-based Service Quality Assurance

클라우드 서비스 도메인에서는 로그 데이터와 시스템 오류 데이터를 분석하여 서비스 안정성과 장애 탐지 능력을 평가하였다. AI는 실시간 로그 분석을 통해 장애 징후를 탐지하고, 인간 운영자는 탐지 결과를 검증하여 복구 절차를 조정하였다.

결과적으로 결함 탐지율은 80%에서 92%로 향상, 오류율은 16%에서 13.7%로 감소하였다. AI의 실시간 분석 능력과 인간의 판단력이 결합되어 시스템 안정성과 복구 효율성이 향상되었다.

1.5 Evaluation Metrics

본 연구는 다음 네 가지 핵심 품질지표를 통해 프레임워크의 성능을 평가하였다.

이러한 지표들은 정량적 효율성(PS, DDR, ACC)과 정성적 신뢰성(ER 감소 효과)을 함께 반영하여, AI의 자동화 속도 향상과 인간 검증의 품질 판단력을 동시에 평가할 수 있는 통합 품질관리 체계를 구성한다.

[Table 3]에서 보여주는 것처럼 제안한 하이브리드 품질관리 프레임워크는 세 도메인 모두에서 AI의 자동화 성능 향상과 함께 인간 전문가의 검증 피드백이 품질 신뢰성을 높이는 데 실질적으로 기여하였다. 특히 EduTech과 클라우드 QA 사례에서는 AI의 편향 수정과 동적 임계값 조정 등 인간 피드백 기반의 학습 메커니즘이 프레임워크의 자기진화(Self-evolving) 구조로 작동하였다.

이러한 결과는 본 프레임워크가 현실 산업 환경에서 재현 가능한 실증적 품질관리 모델로 기능할 수 있음을 보여준다.

- 역할 분담의 명확성: AI는 반복·규칙적 작업, 인간은 창의·윤리적 판단을 전담.
- 상호 피드백 구조: AI 산출물에 대한 인간 피드백이 다시 AI 학습으로 이어짐.
- 정량·정성 통합 평가: 단일 지표 중심의 기존 연구를 넘어, 객관성과 실효성을 동시에 확보.
- 지속적 개선성: 데이터와 피드백 축적을 통한 프레임워크의 자기 진화 가능성.

Table 3. Evaluation Metrics

Metric	Definition	Unit
Defect Detection Rate (DDR)	Ratio of detected defects to total defects	%
Error Rate (ER)	Ratio of false positives to total predictions	%
Processing Speed (PS)	Number of quality checks per hour	checks/hour
Accuracy (ACC)	Ratio of correctly classified predictions	%

2. Evaluation

2.1 Experimental Results

통계분석 결과, 세 도메인 모두에서 $p < 0.05$ 수준의 유의미한 개선이 확인되었다(Table 4 참조). 스마트 제조는 검사 속도가 37.5% 향상되었고, 에듀테크에서는 예측 정확도가 8%p 증가하였다. 클라우드 서비스에서는 장애 탐지율이 12%p 향상되었다. 이는 제안된 프레임워크가 효율성과 신뢰성을 동시에 향상시키는 균형적 품질관리 모델임을 입증한다.

2.2 Discussion

실험한 세 도메인은 산업적 특성은 다르지만, 모두 AI 자동화와 인간 검증의 협업 구조를 기반으로 한다는 공통점을 지닌다. AI는 정량적 효율성을 높이고, 인간은 정성적 신뢰성을 강화함으로써 품질관리의 두 축이 조화를 이루는 균형 구조를 형성하였다. 이러한 하이브리드 품질관리 방식은 기존의 AI 중심 혹은 인간 중심 품질관리 접근의 한계를 극복하고, 속도·일관성·신뢰성을 모두 향상시키는 결과를 가져왔다. 특히 동일한 데이터셋과 실험 환경에서 수행한 비교 실험 결과, 제안된 하이브리드 품질관리 프레임워크는 기존 AI-only 품질관리 방식 대비 정확도, 결함 탐지율, 처리 속도 전반에서 일관된 성능 향상을 보였으며, 이는 인간 전문가의 검증이 AI 자동화의 한계를 효과적으로 보완함을 입증한다.

기존 품질관리 연구는 주로 인간 전문가의 경험적 판단에 의존하거나, AI 기반 자동화 성과에만 초점을 맞추는 경향이 있었다. 이에 반해 본 연구의 프레임워크는 AI 자동화의 효율성과 인간 판단의 윤리적 책임성(Ethical Accountability)을 동시에 고려하여, 품질관리 과정의 균형성과 설명가능성(Explainability)을 제도적으로 확보할 수 있는 구조를 제시한다.

다음 [Table 4]는 기존 품질관리 연구와 제안된 하이브리드 품질관리 프레임워크의 차별성을 정리한 것이다. 이 비교를 통해 제안된 프레임워크는 기존 품질관리 방식에 비해 다음과 같은 혁신성을 갖는 것으로 평가된다.

이러한 비교 결과는 제안된 하이브리드 프레임워크가 정량적 성능 향상뿐 아니라, 조직적·윤리적 품질관리 체계

로 발전할 수 있는 실질적 가능성을 보여준다.

따라서 본 연구는 향후 다양한 산업 영역에서 AI-인간 협업 기반 품질관리의 표준화 모델로 활용될 수 있을 것으로 기대된다. 이에 본 연구는 제안된 프레임워크가 국제 품질표준(ISO/IEC 25010, ISO 9001:2015)과 개념적·구조적으로 어떻게 연계되는지를 다음 절에서 논의한다.

2.3 Relationship with International Quality Standards

본 프레임워크는 특정 산업 표준에 종속되지 않으며, ISO/IEC 25010 및 ISO 9001:2015의 품질속성에 기반하여 설계되었다. ISO/IEC 25010의 주요 속성들 중에 성능 효율성, 신뢰성, 보안성, 유지보수성은 본 프레임워크의 설계 원칙(역할 최적화, 상호 피드백, 다층 평가, 지속 개선)과 대응된다. 또한 ISO 9001의 지속적 개선과 사실 기반 의사결정 원칙은 AI-인간 피드백 루프를 통해 실현된다. 이에 따라 제안된 프레임워크는 품질속성 중심의 도메인 독립적 표준 모델로 확장 가능하며, 산업 전반의 AI-인간 협업 기반 품질관리 체계 구축에 기여할 수 있다. 따라서 본 연구는 다양한 산업 도메인에 적용 가능한 하이브리드 품질관리의 새로운 표준을 제시하며, 실제 현장에서 활용 가능한 실질적 효과를 입증하였다. 또한 본 연구에서는 성능 지표 중심의 평가를 넘어, 인간 전문가 개입이 품질관리 운영에 미치는 영향을 함께 분석하였다. 제안된 프레임워크에서 인간 전문가는 전 과정에 상시 개입하지 않고, AI 자동화 결과 중 신뢰도가 낮은 사례에 대해서만 선택적으로 개입하도록 설계되었다. 이로 인해 반복적인 품질 점검 업무는 AI가 담당하고, 인간 전문가는 핵심 사례에 집중할 수 있어 전반적인 운영 부담이 감소하였다. 이러한 구조는 추가적인 인력 증원 없이도 품질 신뢰성과 운영 효율성을 동시에 확보할 수 있는 비용 대비 효과적인 품질관리 모델임을 시사한다.

V. Conclusions

본 연구는 인공지능(AI)과 인간 전문가의 협업을 기반으로 하는 하이브리드 소프트웨어 품질관리 프레임워크를

Table 4. Quantitative Comparison Results across Three Domains

Domain	Baseline System	Proposed Framework	Improvement
Smart Manufacturing	DDR 89%, PS 72/h	DDR 94%, PS 99/h	+5.6% accuracy, +37.5% speed
EduTech	ACC 82%, ER 18%	ACC 90%, ER 14%	+8.0% accuracy, -22% error
Cloud Service	DDR 80%, ER 16%	DDR 92%, ER 13.7%	+12.0% accuracy, -14% error

제안하였다. 기존의 품질관리 연구가 AI 중심 자동화 혹은 인간 중심 검증 중 하나에 치우친 한계를 보인 반면, 본 프레임워크는 두 접근의 강점을 결합하여 효율성과 신뢰성의 균형을 달성하는 것을 목표로 하였다. 프레임워크는 네 가지 설계 원칙인 역할 최적화, 상호 피드백, 다층 품질평가, 지속 개선을 기반으로 설계되었으며, 국제 품질표준인 ISO/IEC 25010 및 ISO 9001:2015의 품질 속성과 연계된다. 세 산업 도메인에 적용한 실험 결과, 품질검사 처리속도는 평균 38% 향상, 결함 탐지율은 45% 증가, 오류율은 15% 감소하였다. 통계 분석에서도 모든 도메인에서 $p < 0.05$ 수준의 유의미한 개선이 확인되어 AI 자동화의 효율성과 인간 검증의 신뢰성이 상호 보완적으로 작용함을 실증하였다. 이러한 결과는 제안된 프레임워크가 정량적 품질지표와 정성적 품질요소를 동시에 향상시키는 균형적 품질관리 체계로 기능할 수 있음을 보여준다. 향후 연구에서는 다음과 같은 확장 방향이 필요하다. 첫째, 실제 산업 현장에서의 장기적 적용을 통해 프레임워크의 지속적 학습 효과(Continuous Learning Effect)를 검증할 필요가 있다. 둘째, 다양한 AI 모델(예: 대규모 언어모델, 생성형 AI)의 품질관리 프로세스에 본 프레임워크를 접목하여 AI 거버넌스 품질관리 모델로 확장할 수 있다. 셋째, 인간 전문가의 판단과 AI 추천 결과 간의 설명가능성(Explainability)을 강화하는 지표 개발 연구가 향후 품질관리 신뢰도 향상에 기여할 것으로 기대된다.

ACKNOWLEDGEMENT

The present research has been conducted by the Research Grant of Seoil University.

REFERENCES

- [1] Pressman, R. S., & Maxim, B. R., "Software Engineering: A Practitioner's Approach" (9th Edition). McGraw-Hill Education, 2020.
- [2] ISO/IEC 25010:2011. "Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models." International Organization for Standardization, 2011.
- [3] Juran, J. M., & Godfrey, A. B., "Juran's Quality Handbook" (5th Edition). McGraw-Hill, 1999.
- [4] IBM, "AI in Software Development", 2024, <https://www.ibm.com/kr-ko/think/topics/ai-in-software-development>.

- [5] NIA, "2025 AI Data Quality Management Guidelines v3.5." 2025, https://www.nia.or.kr/site/nia_kor/ex/bbs/View.do?cblIdx=26537&bcIdx=28106&parentSeq=28106.
- [6] Wu, S., et al., "Human-generative AI collaboration enhances task performance and quality." *Nature Scientific Reports*, 15, Article 98385, 2025, <https://www.nature.com/articles/s41598-025-98385-2>.
- [7] Zhang, Y., et al., "Quality of human-GenAI collaboration and its driving factors." *Information & Management*, 62(4), Article 3140, 2025, <https://www.sciencedirect.com/science/article/pii/S0306457325003140>.
- [8] Fragiadakis, G., Diou, C., Kousiouris, G., & Nikolaidou, M., "Evaluating Human-AI Collaboration: A Review and Methodological Framework." *arXiv:2407.19098 [cs.HC]*, 2024.
- [9] Dignum, V., "Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way." Springer, 2019.
- [10] Eunsook Cho, "Development of Metrics to Measure Reusability Quality of AlaaS", *Journal of The Korea Society of Computer and Information*, Vol. 28 No. 12, pp. 147-153, December 2023, <https://doi.org/10.9708/jksci.2023.28.12.147>
- [11] Bird, C., et al., "Human-AI, Collaboration in Software Engineering: Lessons from Industry Practice." *Proceedings of the ACM on Software Engineering*, 3(2), pp.45-67. <https://dl.acm.org/doi/10.1145/3643690.3648236>, 2024.
- [12] Han, J., et al., "Developer-LLM Conversations: Empirical Study of Human-AI Collaboration in Code Generation." *Hanbit Media Research Reports*, 2025. <https://m.hanbit.co.kr/channel/view.html?cmscode=CMS6360971093> 2025.
- [13] ISO 9001:2015. "Quality management systems-Requirements." International Organization for Standardization, 2015.
- [14] Sommerville, I., "Software Engineering" (10th Edition). Pearson, 2016.

Authors



Eun-Sook Cho received the B.S. degree in Computer Science from DongEui University, Korea in 1993. She received the M.S and Ph.D degree in Computer Science from SoongSil University, Korea, in 1996 and

2000, respectively. Dr. Cho joined the faculty of the Department of Software Engineering at Seoil University, Seoul, Korea, in 2005. She is currently a Professor in the Department of Software Engineering, Seoil University. She is interested in framework modeling and development, Software Quality, AI, BigData, Service-Oriented Computing, and IoT Applications.